How Opaning Schools Influence COVID Infections – Empirical Evidence from Czech Republic

April 27, 2021

Main findings

- Opening of schools increases the infection numbers of students
- Our data are not variable enough to assess impact of orders to wear masks
- Data do not support the hypothesis that the increase of reported cases after opening schools is due to more intense testing in schools
- Neither they support the hypothesis that, when being out of schools, children are infected elsewhere
- Opening of closed schools on April 12th, 2021 in CR without restrictions other than masks would bring additional approx 2500 weekly cases, CI_{95} (1765,3051)

Methods

Denote $X_t^1, \dots, X_t^5$ the numbers of newly infected individuals from age cohorts $3-5$, $6-10$, $11-14$, $15-18$, all remaining, respectively, where t is time in weeks. Denote $S_t^1, \dots, S_t^4$ the percentage of opening kindergarten/first grade/second grade/secondary school classes, and denote $M_t^1, \dots, M_t^4$ zero-one variables indicating whether wearing in classes masks was ordered.

Our goal is to examine whether opening of schools contributes to the epidemic and whether masks play role in schools. To this end, each cohort i, we assume that the expected number of newly infected children outside schools is Compound Poisson with mean

$$\beta_t^i X_{t-1}, \qquad X_{t-1} = \sum_{j=1}^{5} X_{t-1}^j, \qquad \beta_t = R_t \beta,$$

where β is an unknown constant and R_t the reproduction number.

Further, we hypothesize that the number of children of the i-th cohort, infected at school, is Compound Poisson with mean

$$\gamma_t^i(1 - M_{t-1}^i\mu)S_{t-1}^i X_{t-1}^i \qquad \gamma_t^i = r_t\gamma^i$$

where μ^i is an unknown efficiency of masks, γ^i is an unknown constant and $r_t = R_t/C_{t-2}$ is the the basic reproduction number, C_τ is the contact restriction at τ.

The logic behind our model is following: without schools, the cohort of interest is infected "as usual" which means that the rate of their infection is proportional to the reproduction number and the number of infected in the whole population (a proxy for which up to constant is the daily number of infected is). In schools, the situation is similar with the difference that we explicitly model the contact intensity therein (by variables S and M and constants γ and δ), so the overall contact restriction does not play role; therefore, we assume the infection to be proportional to r rather than R.

Summed up, we have

$$\begin{aligned} X_t^i &= \beta_t^i X_{t-1} + \gamma_t^i(1 - M_{t-1}^i\mu)S_{t-1}^i X_{t-1}^i + \epsilon_t^i \\ &= \beta^i R_t X_{t-1} + \gamma^i r_t S_{t-1}^i X_{t-1}^i + \delta^i r_t M_{t-1}^i S_{t-1}^i X_{t-1}^i + \epsilon_t^i, \qquad \delta^i = \gamma^i\mu. \end{aligned}$$

where $\mathbb{E}\epsilon_t^i = 0$ and $\operatorname{var}(\epsilon_t^i)$ is, up to a constant, equal to the r.h.s. minus the residuum.

Finally, we assume that the observed numbers $Y_t^1, \dots, Y_t^5$ of the infected are proportional to the actual ones, namely $Y_t^i = cX_t^i + e_t^i$ where c is an unknown constant and e_t^t are centered with $\mathrm{var(e_t^i)} \sim X_t^i$. This gives

$$Y_t^i = \beta^i I_t + \gamma^i U_t^i + \delta^i V_t^i + \eta_t^i, \qquad I_t = R_t Y_{t-1}, \qquad U_t^i = r_t S_{t-1}^i Y_{t-1}^i, \qquad V_t^i = r_t M_{t-1}^i S_{t-1}^i Y_{t-1}^i$$

$$\eta_t^i = \frac{\epsilon_t^i}{c} + r_t\left[\beta^i \sum_{j=1}^k e_t^j + (\gamma^i S_{t-1}^i + \delta^i M_{t-1}^i S_{t-1}^i)e_t^i\right] \qquad (1)$$

Note that $\mathbb{E}\eta_t^i = 0$ and $\operatorname{var}(\eta_i^i) = \sum_j d_t^{i,j} X_t^j$ for some $d_{i,t}^t$. As, in practice, the ratios of X_t^i does not vary much in time, we approximate $\operatorname{var}(\eta_i^i) \doteq dX_t \doteq \frac{d}{c} I_t$ where d is unknown parameter.

Our hypotheses are

$$H_0^i : \gamma^i = 0 \text{ against } H_1^i : \gamma^i > 0$$

(schools do not/do have influence) and

$$\tilde{H}_0^i : \delta^i = 0 \text{ against } \tilde{H}_1^i : \delta^i < 0$$

(masks do not / do have influence).

We use WLS to estimate coefficients in (1), this, however, means OLS after dividing the equation by $\sqrt{I_t}$ for each t, and test the hypotheses by t-tests for each i.

Data

The values Y are taken from the data issued by the Czech Ministry of Health, namely the anonymized person-level dataset (osoby.htm) each record of which includes date of reporting the infection and the age.

The reproduction number is estimated as $R_\tau = \frac{T_\tau - T_{t-7}}{T_{\tau-5} - T_{t-12}}$ where T is the accumulated number of reported infections and τ is time in days. The weekly values are adjusted by exponential smoothing with factor $\alpha = 0.3$. The contact reduction is taken from survey \cite{paqcovid} and adjusted by exponential smoothing with factor $\alpha = 0.3$.

The values of S and M are estimated from publicly available data (see the electronic supplement). Only values starting from April 2020 are taken into account, as we regard the previous data as noisy and unreliable. Last values are from the first week of April 2021. The observations corresponding to summer holidays are excluded as well the observations corresponding to weeks where the rotational schooling took place (see the electronic supplement). The following table shows numbers of observations at our disposal

	Kindergartens	First Grade	Second Grade	Secondary
Observaions	44	43	44	44
Opened	34	26	13	24
... with masks	21	22	9	23
... without masks	13	4	4	1

Results

Estimation of equations (1) gave ambiguous results: For kindergartens, all β, γ and δ came out significantly, but with $\delta > 0 > \gamma$. For the first grade, only β came out significantly. For the remaining two cohorts, all the coefficients came out significantly, consistently with our model ($\gamma > 0 > \delta$, $|\delta| < \gamma$). First two ambiguous results are possibly most likely by colinearity - the correlations of U^i and V^i is greater than 0.9 for all i, which casts shoadows also on the results of the latter cohorts

Models with excluded δ, i.e. not measuing the effect of masks, all came out significantly, exhibiting relatively good fit in terms of centered R^2 :

	kindergartens	first grade	second grade	secondary
β	$0.011(0.0014)^{***}$	$0.030(0.0023)^{***}$	$0.032(0.0013)^{***}$	$0.026(0.0014)^{***}$
γ	$0.36(0.05)^{***}$	$0.60(0.14)^{***}$	$0.22(0.10)^{**}$	$0.50(0.07)^{***}$
cent. R^2	0.81	0.76	0.75	0.83

We also considered an alternative explanation of potential increase of infections when schools are open, saying that the increase of cases is not caused by being in school (i.e. $\gamma^i = 0$), but by more intense testing when children are at school, i.e. $Y_t^i = (c + S_t^i \phi^i) X_t^i$ for some ϕ^i. Given this addition, we would have

$$Y_t^i = \beta^i I_t + \phi^i W_t^i + \gamma^i U_t^i + \tilde{\eta}_t, \qquad \psi^i = \beta \phi^i, \qquad W_t^i = I_t S_{t-1} \tag{2}$$

and test

$$\hat{H}_0^i : \psi^i = 0 \text{ against } \hat{H}_1^i : \psi^i > 0$$

For all the cohorts, however, ψ^i came out insignificant while, for kindergartens, it came out significant negative.

Finally, to avoid dependence on our guesses of degree of openness, we re-estimated the equations using excluding records with $0 < M_t^i < 0.9$ with the following result

	kindergartens	first grade	second grade	secondary
β	0.015(0.0017)***	0.026(0.0018)***	0.031(0.0018)***	0.025(0.0017)***
γ	0.32(0.059)***	0.40(0.12)***	0.28(0.12)**	0.53(0.068)***
cent. R^2	0.81	0.80	0.66	0.86

Using these data, we can estimate that opening of schools on April 12, 2021 without restrictions other that wearing masks in class would bring the following increases in weekly new cases

	kindergartens	first grade	second grade	secondary	all
mean	314	689	514	891	2408
lower 95%	201	284	82	667	1765
upper 95%	427	1093	946	1115	3051

Discussion

Inconsistencies in kinderkgartens

Covers the hypothesis that children are infected else

Sinsitivity to our estimates of S

Appendix

TBD